DOCUMENT RESUME

ED 208 042 TM 810 741

AUTHOR Halpin, Gerald; And Others
TITLE Research on Writing: A Search for Objective Measures Related to Holistically Scored Essays.
PUB DATE Mar 81
NOTE 13p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Philadelphia, PA, March, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Correlation; *Essay Tests; Higher Education; Holistic Evaluation; *Multiple Choice Tests; Testing Problems; *Test Validity; Writing (Composition); Writing Research; *Writing Skills
IDENTIFIERS Missouri College English Test

ABSTRACT

Controversy has characterized the research on writing. On the one side are those who state that, to evaluate achievement in writing, evaluate the writing of students. On the other side are those pointing out the problems associated with direct measures of writing such as low reliability and high cost in terms of time and often dollars. The purpose of this study was to determine if select objectively measured indirect indicators of writing ability do account for scores obtained on holistically scored direct measures of writing ability. The indirect measure was the Missouri College English Test, composed of 90 multiple-choice items assessing punctuation, capitalization, grammar, spelling, sentence style and structure, and paragraph organization. The direct measure was an essay on a general topic, to be written in 30 minutes. Bivariate and multiple correlational analyses were conducted between each subtest of the Missouri test and the essay. All six independent measures together accounted for 26 percent of the variance in the essay scores. These results might make questionable the sole use of indirect measures for making decisions about competence in written English. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED208042

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

RESEARCH ON WRITING: A SEARCH FOR OBJECTIVE MEASURES RELATED TO HOLISTICALLY SCORED ESSAYS

Gerald Halpin, Glennelle Halpin,
and Barbara Brown Schaer
Auburn University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Halpin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 741

Paper presented at the meeting of the Eastern Educational Research Association, Philadelphia, March 1981.

RESEARCH ON WRITING: A SEARCH FOR OBJECTIVE MEASURES RELATED TO HOLISTICALLY SCORED ESSAYS

Years ago, Diederich (1946) said that "the essay is unquestionably a valid test of ability to write, for it is an instance, a sample, of the very ability that one is attempting to measure. There is no more direct evidence of ability to write" (p. 584). Eley (1955) added that "an adequate essay test of writing is valid by definition; that is to say, it has face validity since it requires the candidate to perform the actual behavior which is being measured" (p. 11). Brown (cited in Palmer, 1961) summed up arguments such as these and others favoring the essay by saing: "Everybody agrees that writing should be tested by writing" (p. 472). However, obtaining reliable estimates of writing ability via the essay continues to be an expensive, time-consuming process (Coffman, 1971; Lutz, Note 1, Note 2). That writing samples must be obtained on multiple occasions and scored by several raters to obtain very reliable estimates of students' true abilities has been demonstrated in a number of studies (Braddock, Lloyd-Jones, & Schoer, 1963, Coffman, 1966, Godshalk, Swineford, & Coffman, 1966; Llabre, 1978; McColly, 1970).

Indirect measures of writing ability have arisen as an alternative to direct assessments requiring that essays be written and scored. These indirect assessments require no writing at all but instead require that examinees respond to related questions cast in a multiple-choice format. Indirect measures usually have greater reliability and consequently tend to be favored by those concerned with quality measurement. Arguing for indirect assesment, Noyes, Sale, and Stalnaker (1945) suggested that a student assessed by means of an essay is somewhat in the position of a gambler who risks all on a single throw of the dice, while a multiple-choice test allows many throws: "The good

candidate who errs on a few of these (multiple-choice items) has plenty of opportunity to redeem himself; a mistake on one item does not affect any other item. In writing a theme, however, the candidate who makes a false start almost inevitably involves his whole theme in difficulties even though he may be, generally speaking, a good writer" (p. 9). Objecting to the assessment of writing skill via multiple-choice tests, though, Braddock, Lloyd-Jones, and Schoer (1963) wrote that "not only do they (multiple-choice tests) not require the examinee to perform the actual behavior being measured--he does no actual writing; but these tests also make little or no attempt to measure the 'larger elements' of composition, even indirectly" (p. 42).

It is apparent from the literature that controversy has characterized the research on writing. On the one side are those agreeing with McCaig (1977) who stated that to evaluate achievement in writing, evaluate the writing of students. On the other side are those pointing out the problems associated with direct measures of writing such as low reliability and high cost in terms of time and often dollars. Due to such difficulties with direct assessment of writing, many practitioners have consequently opted to use indirect measures. Akeju (1972), for example, concluded that the reliability of the West African General Certificate Education examinations in English composition was inadequate and suggested the use of multiple-choice tests. The College Board has used objective (i.e., multiple-choice) English tests for years (Palmer, 1961). Before objective tests of English competence are used in lieu of writing samples, however, it is important that they be demonstratively related to essays. They should indeed, considering the charges of Braddock, Lloyd-Jones, and Shoer (1963), measure the various elements of composition.

The purpose of this study was to determine if select objectively measured indirect indicators of writing ability do account for scores obtained on holistically scored direct measures of writing ability. Specifically investigated were the bivariate and multivariate relationships between the independent or predictor variables of punctuation, capitalization, grammar, spelling, sentence style and structure, and paragraph organization as measured by the Missouri College English Test and the dependent or criterion variable of essay score obtained via the holistic scoring procedure.

Procedure

Participants in this study were 172 sophomores and juniors who were applicants for admission to teacher education at a large state university who had to pass an English proficiency examination prior to acceptance into the teacher education program. Since participation in this English proficiency testing program is required, all applicants were examined thereby preventing a selection problem due to non-cooperation. Both direct and indirect assessments were made.

The indirect measure of writing ability was the Missouri College English Test (Callis & Johnson, 1965), an objective measure composed of 90 multiple-choice items assessing competence in six areas: punctuation, capitalization, grammar, spelling, sentence style and structure, and paragraph organization. It was administered to the subjects in a monitored, test-like condition in a large university auditorium at a time designated for the testing. Student responses were electronically scored for a total score and scores on the six subtests of punctuation, capitalization, grammar, spelling, sentence style and structure, and paragraph organization. The resulting raw scores were used in the analyses.

Internal consistency alpha reliability coefficients were computed for the total and for each subtest. As shown in Table 1, the reliability estimates

---

Insert Table 1 About Here

---

ranged from .55 to .73 for the subtests and was .80 for the total. These reliabilities were generally low but sufficient for research purposes.

The direct measure of writing ability was an essay which the subjects wrote after completing the Missouri English test. The topic they were given to write about was very general (e.g., friendship), and they were instructed to narrow the topic down and write the essay in 30 minutes.

Three faculty members who had professional preparation for and experience in teaching English and/or English education at the University level were chosen to evaluate the writing samples. The holistic method of scoring was utilized. This process is based upon a generalized impression or global quality of a paper. Based upon the recommendation of Coffman (1971) a 10-point scale was chosen.

The three judges were brought together and given a 1-hour training session on scoring procedure. Next, they rated seven sample essays which were not part of the 172 essays being used as data in this study. An average interrater reliability of .82 was obtained for the seven sample papers using the analysis of variance method recommended by Ebel (1979). The average ratings assigned by the three raters were also very close: 6.14, 6.14, and 7.00, respectively. With this kind of agreement, the raters proceeded to rate the 172 essays which were the data source for this study. After all 172 papers were scored by all three raters, the average interrater reliability was again computed. The

resulting coefficient was .86 (see Table 1). The average ratings given by the three raters were also most congruent: 5.81, 6.20, and 6.66, respectively.

Data Analyses and Results

Bivariate and multiple correlational analyses were conducted between each subtest of the Missouri test and the essay (score = average of the three judges' ratings) using stepwise regression procedures of the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975). A Pearson r was also computed between the Missouri test total and the average essay ratings. (See Table 1 for raw score means and standard deviations.)

Results of the correlational analyses are shown in Table 2. As can be seen in this table, the bivariate r's between each subtest on the Missouri test and

---

Insert Table 2 About Here

---

the essay ranged from .25 to .40 ($p < .01$). The multiple R using all six Missouri subtests was .51 ($p < .01$) which was very close to the bivariate correlation between the total Missouri (sum of subtest scores) and the essay, r = .49, $p < .01$.

When the bivariate correlations were corrected for unreliability of the indirect measures, the resulting theoretical correlations ranged from .31 to .54 (see Table 2). Although a reliability estimate of the essay was not available (the reported interrater reliability is actually a measure of objectivity of scoring), the observed reliability for the essay in this study would be much less than unity (1) if essay reliabilities from other studies are generalized. With such being the case, the true relationship between the indirect and direct

measures in this study would have further increased if a correction for attenuation (unreliability) in the essay measure could have also been made.

Discussion

This study suggests, as have previous studies (cf. Breland & Gaynor, 1979; Hogan & Mishler, 1980; Crocker, Ondrasik, & Lamme, Note 3), that indirect and direct measures of writing ability tend to tap similar skills. All six of the objective measures considered accounted for a significant portion of the variance in writing performance in this study. From a bivariate perspective these predictors accounted for 5% to 16% of the variability in writing performance. All six independent measures together accounted for 26% of the variance which is only 2% more variance accounted for than was obtained by simply using a total Missouri score.

From the opposite perspective, however, the six indirect objective measures in combination leave about 74% of the variability in writing unexplained. Each of the indirect measures explained some of the individual differences in the direct measure of writing, but the correlations, although significant, were small. From an educational perspective, the fact that these indirect measures failed to account for a larger percentage of the variability in the writing sample might make questionable the sole use of such measures for making decisions about competence in written English. Better educational decisions might be made using a combination of both direct and indirect measures of writing proficiency, particularly since the raters in this study seemed to be able to tap a broad spectrum of writing abilities in their holistic scoring instead of allowing specific variables such as grammar and spelling that are measurable objectively to dominate their ratings.

Reference Notes

1. Lutz, W. D. Statewide testing in New Jersey. Paper presented at the meeting of the Modern Language Association, San Francisco, December 1979.

2. Lutz, W. D. How to read 55,000 essays a year, and love it. Paper presented at the meeting of the Conference on College Composition and Communication, Washington, D. C., March 1980.

3. Crocker, L., Ondrasik, T., & Lamme, L. Predicting children's writing performance from standardized achievement tests. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, April 1979.

References

Akeju, S. A. The reliability of General Certificate of Education Examination English composition papers in West Africa. Journal of Educational Measurement, 1972, 9, 175-180.

Braddock, R., Lloyd-Jones, R., & Schoer, L. Research in written composition. Champaign, Illinois: National Council of Teachers of English, 1963.

Breland, H. M., & Gaynor, J. L. A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, 1979, 16, 119-128.

Callis, R., & Johnson, W. Missouri College English Test. New York: Harcourt, Brace & World, 1965.

Coffman, W. E. On the validity of essay tests of achievement. Journal of Educational Measurement, 1966, 3, 151-156.

Coffman, W. E. Essay examinations. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D. C.: American Council on Education, 1971.

Diedrich, P. B. Measurement of skill in writing. School Review, 1946, 54, 584-592.

Ebel, R. L. Essentials of educational measurement (2nd ed.). Englewood Cliffs, New Jersey: Prentice-Hall, 1979.

Eley, E. G. Should the General Composition Test be continued? The test satisfies an educational need. College Board Review, 1955, 25, 10-13.

Godshalk, F. I., Swineford, R., & Coffman, W. E. The measurement of writing ability. Princeton, New Jersey: The College Board, 1966.

Hogan, T. P., & Mishler, C. Relationships between essay tests and objective tests of language skills for elementary school students. Journal of Educational Measurement, 1980, 17, 219-227.

Llabre, M. M. An application of generalizability theory to the assessment of writing ability. Unpublished doctoral dissertation, University of Florida, 1978.

McCaig, R. A. What your director of instruction needs to know about standardized English tests. Language Arts, 1977, 54, 491-495.

McColly, W. What does educational research say about the judging of writing ability? Journal of Educational Research, 1970, 64, 148-156.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill, 1975.

Noyes, E. S., Sale, W. M., & Stalnaker, J. M. Report on the first six tests in English composition. New York: The College Board, 1945.

Palmer, O. Sense or nonsense? The objective testing of English composition. The English Journal, 1961, 50, 314-320.

Table 1

Means, Standard Deviations, and Reliability Coefficients for the Missouri College English Test and the Essay Test

Test	$\bar{X}$	SD	Reliability
Missouri College English Test			
Punctuation	7.58	2.10	.55*
Capitalization	8.03	2.52	.67*
Grammar	10.70	2.76	.59*
Spelling	9.76	2.58	.59*
Sentence Style and Structure	6.33	2.03	.56*
Paragraph Organization	14.03	5.80	.73*
TOTAL	59.11	11.80	.80*
Essay	18.61	4.10	.86**

Note. * = alpha internal consistency reliability
** = average interrater reliability

Table 2

Multiple Correlations, R Square, Increase in R Square, Pearson Correlations, Corrected Correlations, and Standardized Beta Weights for Direct and Indirect Measures of Writing

Variable	Multiple R	R Square	Increase in R Square	Pearson r	Corrected r	Standardized Beta
Punctuation	.40*	.16	.16	.40*	.54	.20*
Spelling	.46*	.21	.05	.39*	.52	.17*
Paragraph Organization	.49*	.24	.03	.26*	.30	.14*
Sentence Style and Structure	.50*	.25	.01	.35*	.47	.11
Grammar	.51*	.26	.01	.35*	.46	.10
Capitalization	.51*	.26	.00	.25*	.31	.04
Missouri Total				.49*	.55	

* $p < .01$.